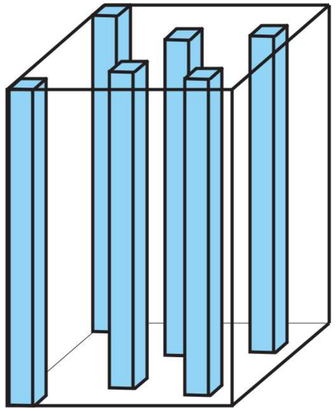
SAND2020-6785C

Practical Leverage Score-Based Sampling for Low-Rank Tensor Decompositions

Brett Larsen
Stanford University
DOE Computational Science Graduate Fellow

Joint work with:
Tamara G. Kolda
Sandia National Laboratories

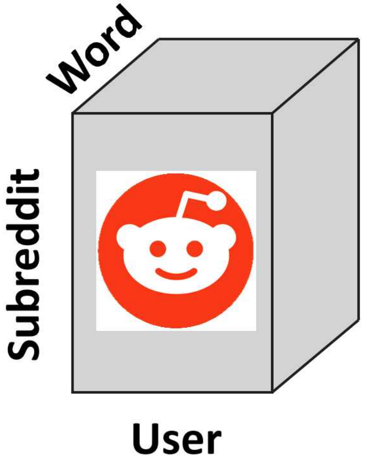


Tamara G. Kolda

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Example Sparse Multiway Data: Reddit

- Reddit is an American social news aggregator, web content rating, and discussion website
  - A “subreddit” is a discussion forum on a particular topic
- Tensor obtained from frost.io (http://frostt.io/tensors/reddit-2015/)
  - Build from reddit comments posted in the year 2015
  - Users and words with less than 5 entries have been removed



Reddit Tensor

8 million users
200 thousand subreddits
8 million words

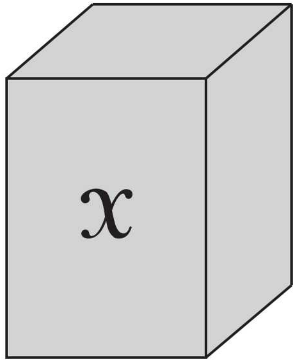
4.7 billion non-zeros ($10^{-8}\%$)
106 gigabytes

$x(i,j,k) =$ log (1 + the number of times user i used word j in subreddit k)

Used a rank $\boldsymbol{r = 25}$ decompsition

Smith et al (2017). “FROSTT: The Formidable Open Repository of Sparse Tensors and Tools”

CP Decomposition into Rank-1 Components

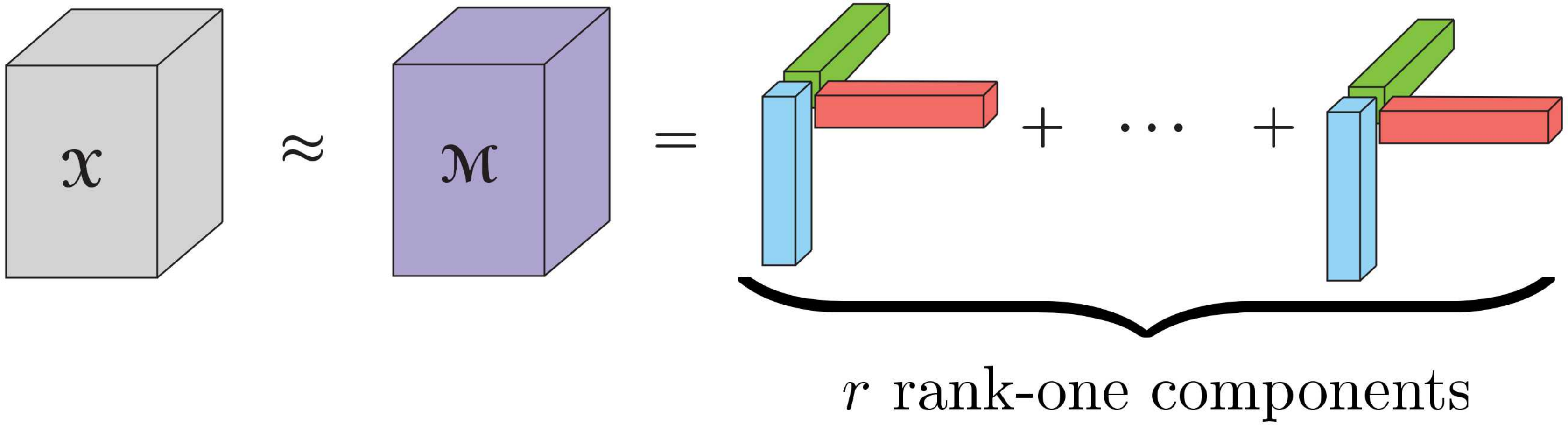


Tensor Properties

Order: 3 | $\mathcal{X}$ is a 3-way tensor.

Length of each dimension: n_1, n_2, n_3

$$\mathcal{X} \text{ is } n_1 \times n_2 \times n_3$$



Rank r Decomposition

Tensor Decomposition Identifies Factors

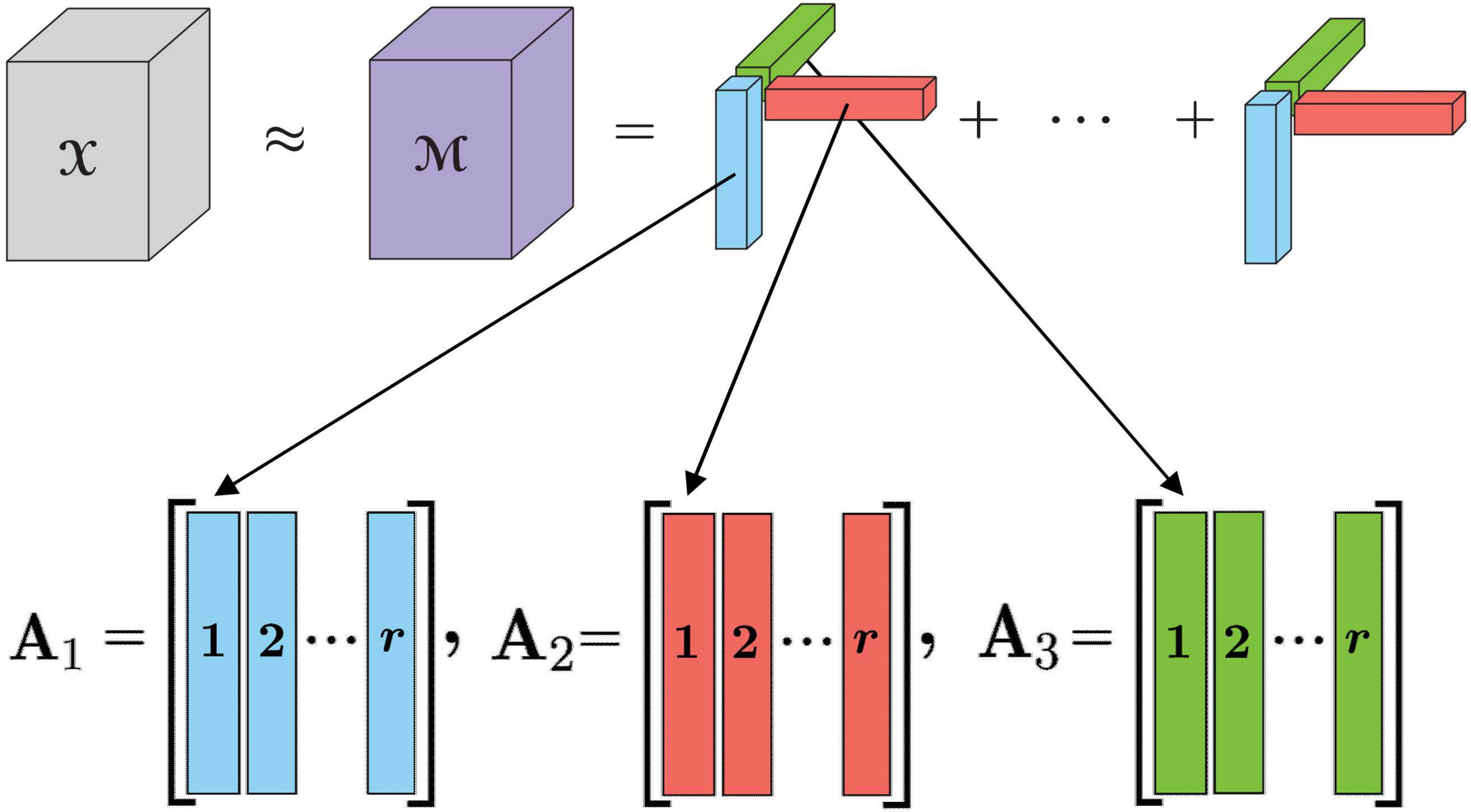
Multi-Index Notation

$$x_i \equiv \mathfrak{X}(i_1, i_2, i_3)$$

Ω: Set of all multi-indices

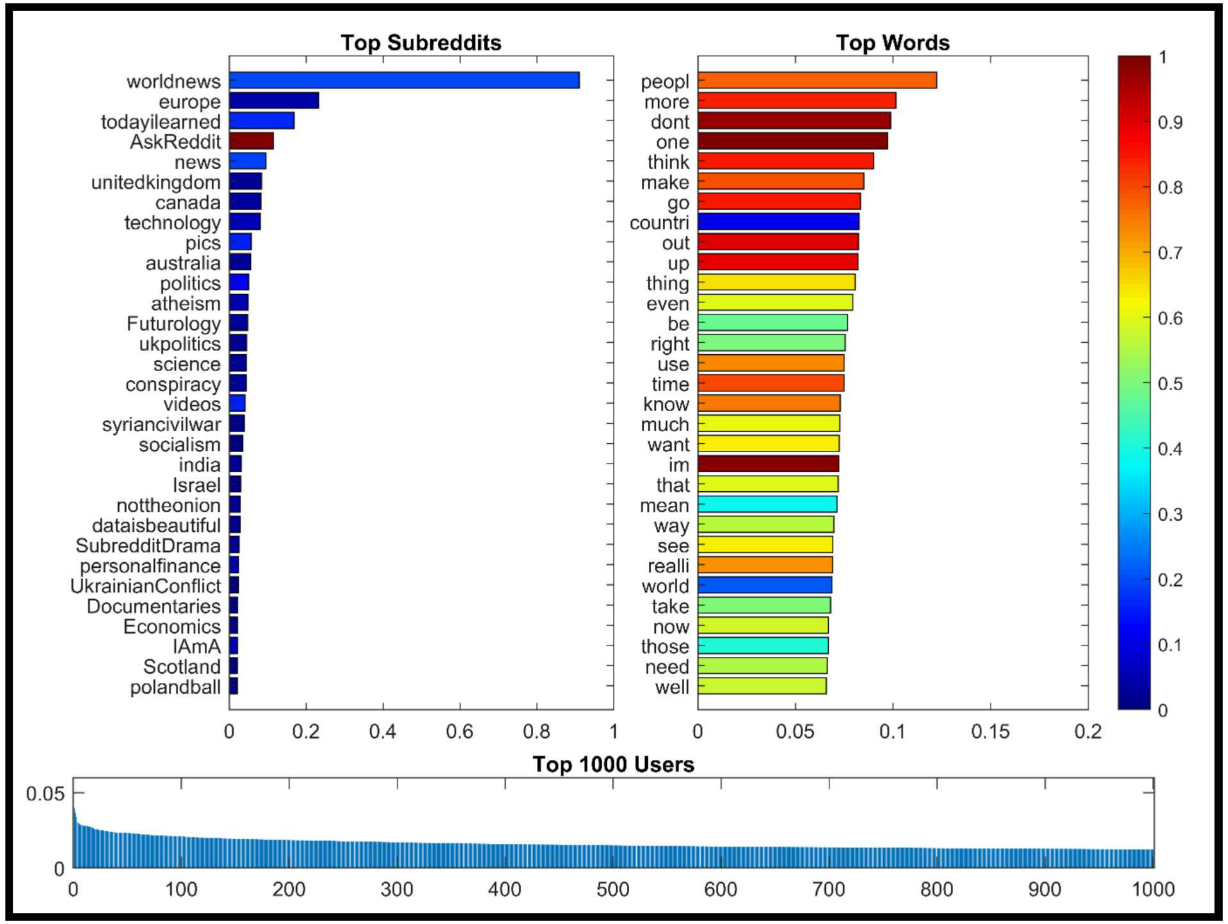
CP Decomposition

$$x_i \approx \sum_{j=1}^{r} \mathbf{A}_1(i_1, j) \ldots \mathbf{A}_3(i_3, j)$$

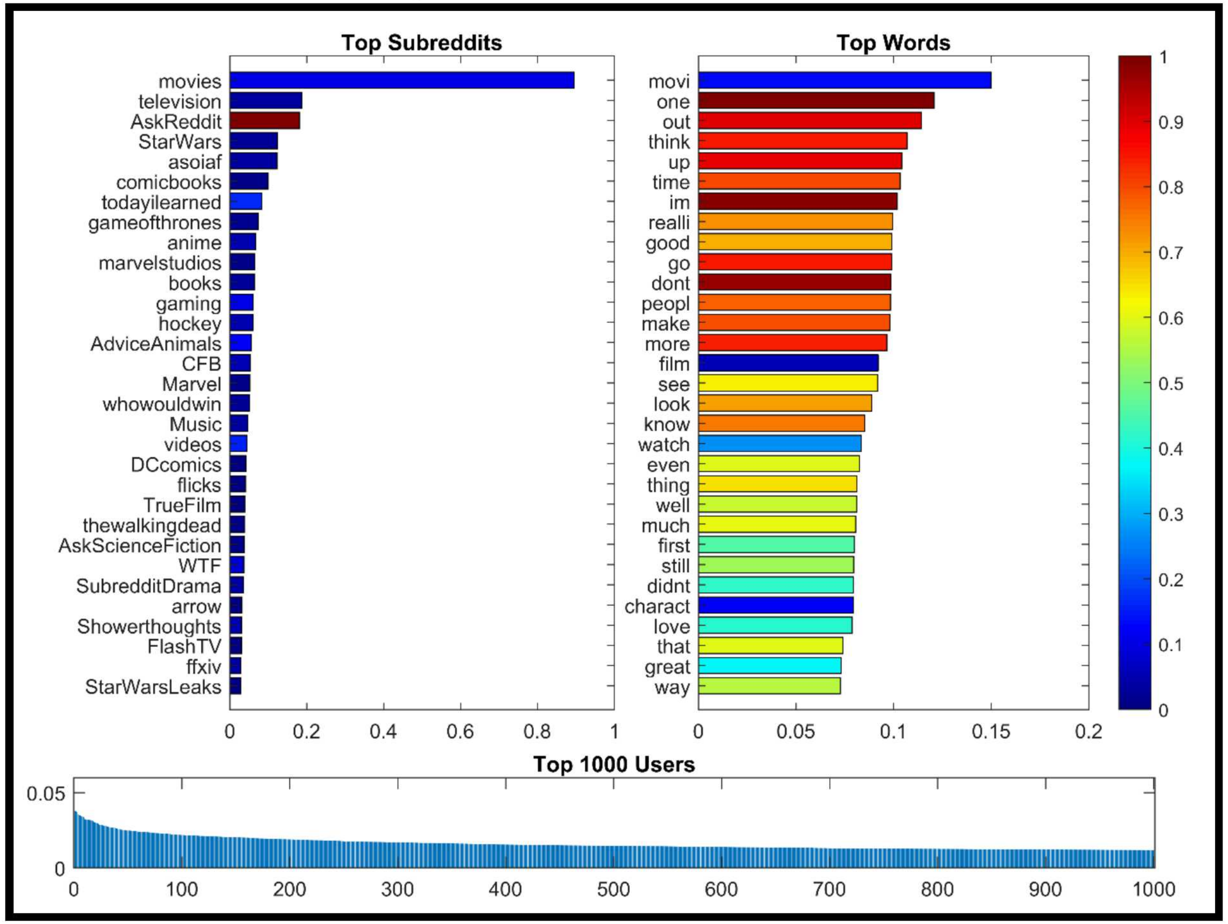


Example Reddit Components Include Rare Words Apropos to High-Scoring Subreddits

Component #6: International News

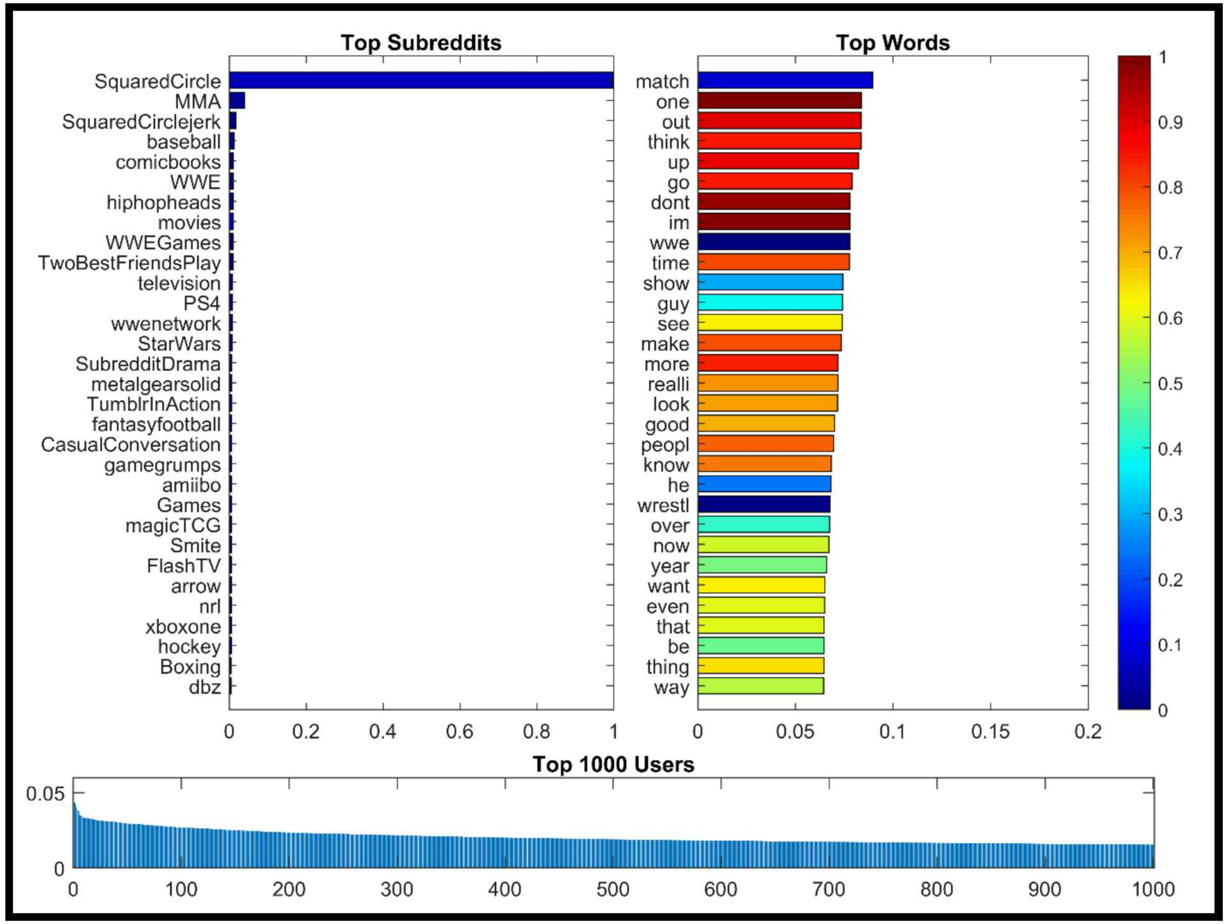


Component #19: Movies & TV

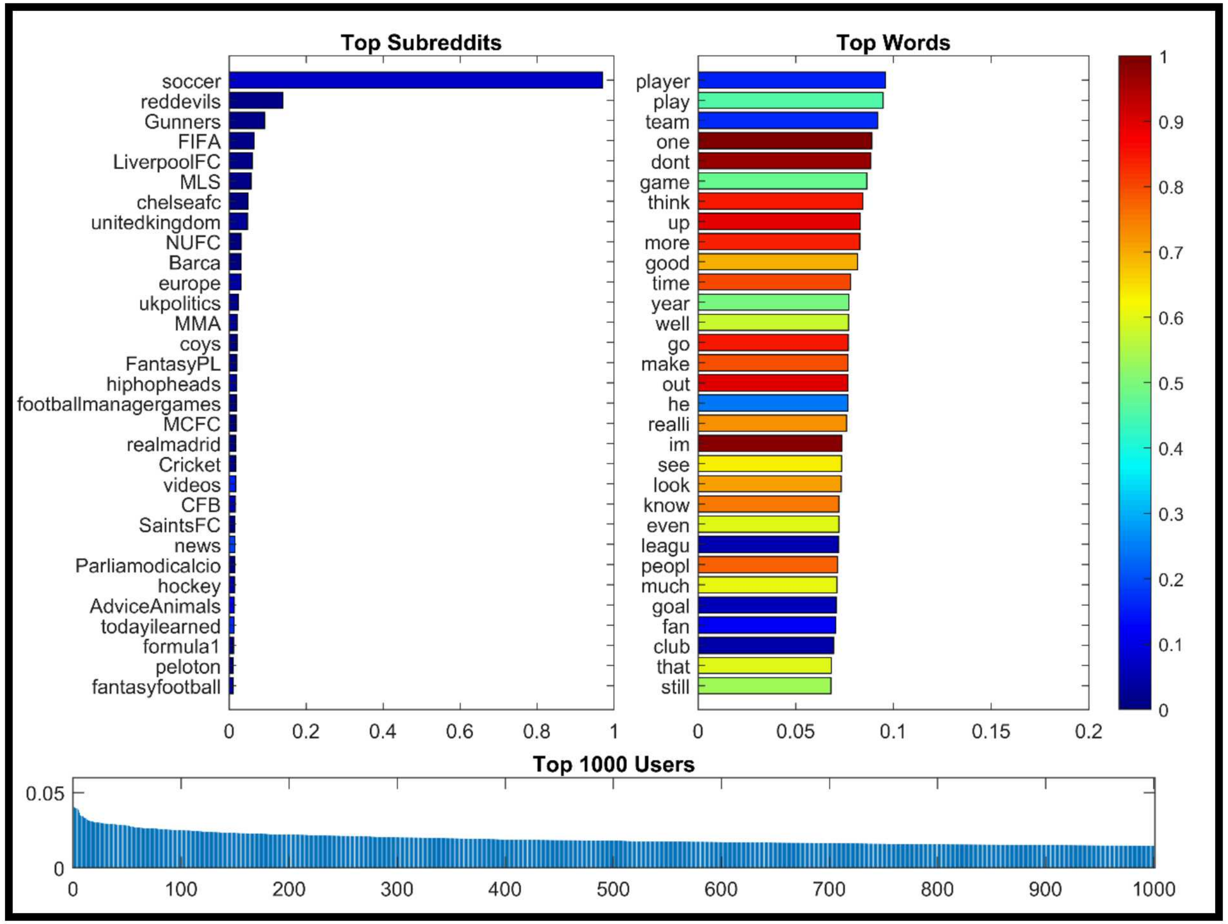


Example Reddit Components Include Rare Words Apropos to High-Scoring Subreddits

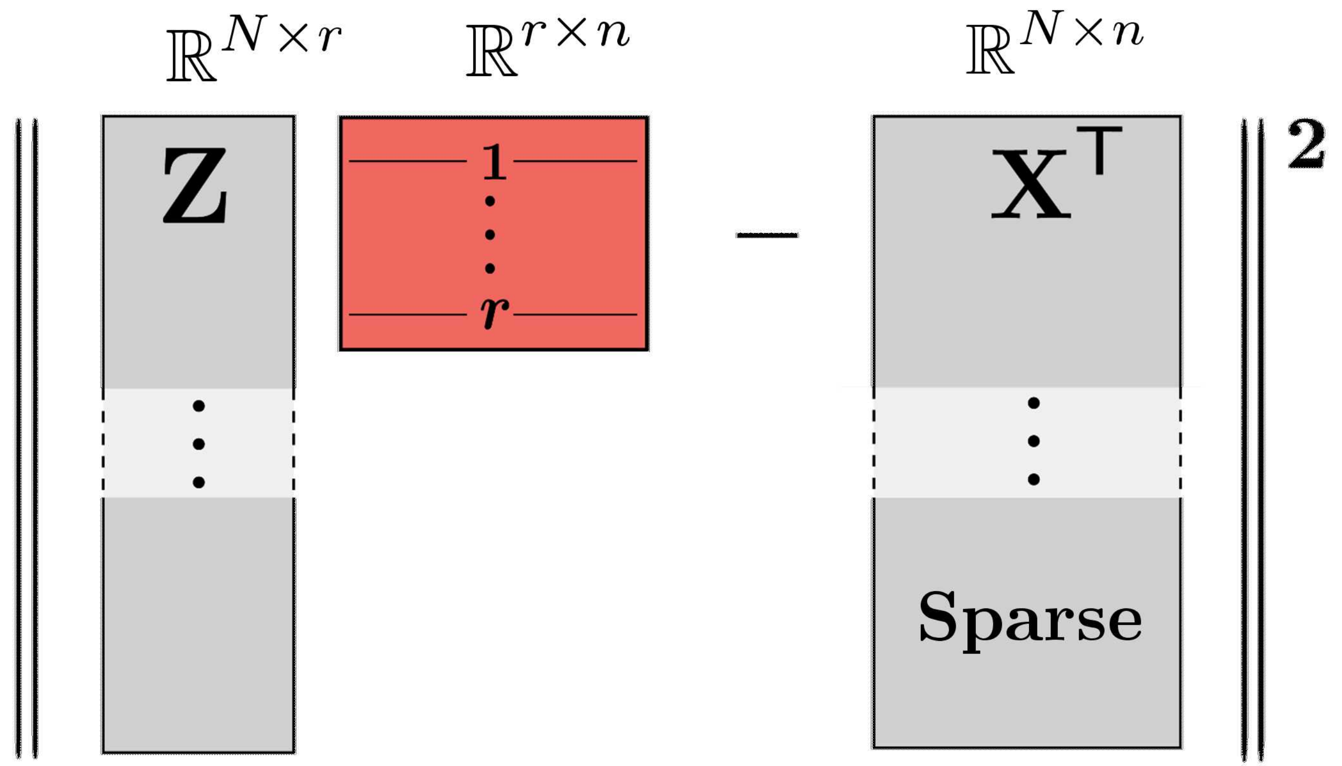
Component #15: Wrestling



Component #18: Soccer



Prototypical CP Least Squares Problem has Khatri-Rao Product (KRP) Structure



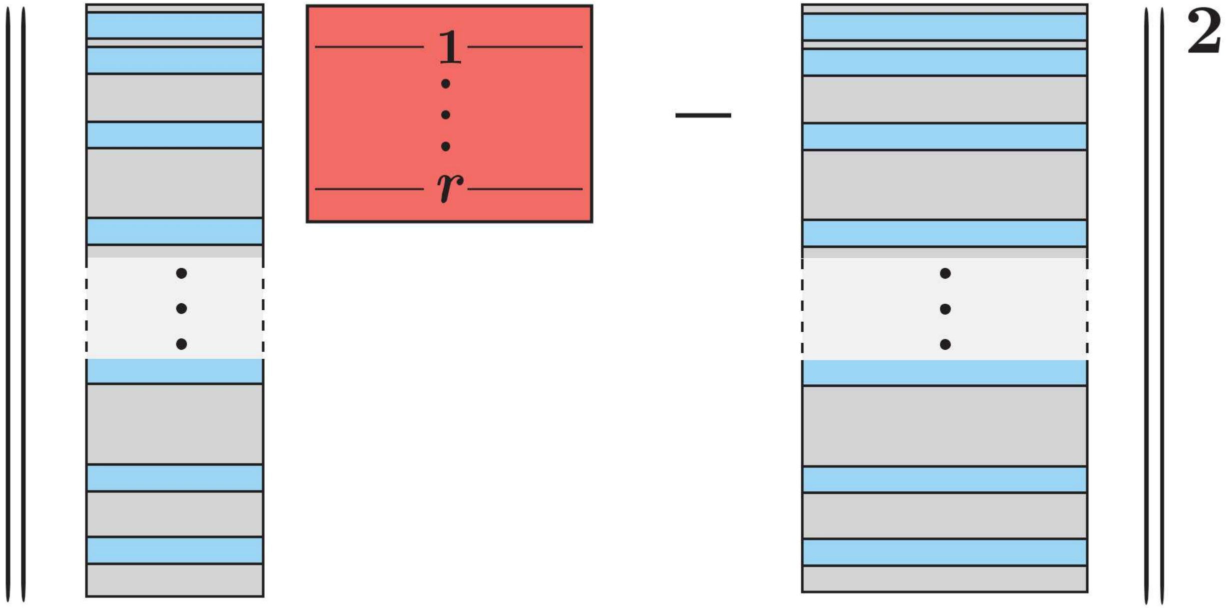
$$\min_{\mathbf{A}} \|\mathbf{Z} \mathbf{A}^{\mathsf{T}} - \mathbf{X}^{\mathsf{T}}\|^2$$

Alternate between factor matrices.

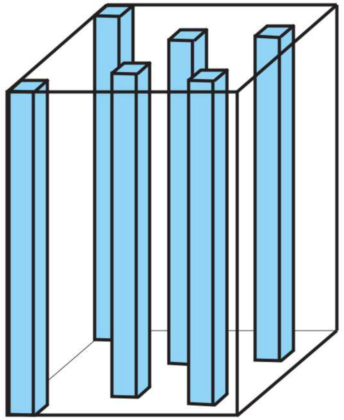
- Subproblem for CP-ALS
  - Solve $d+1$ such subproblems per outer iteration
  - Size n = one of the tensor dimensions
- Matrix $\mathbf{Z}$ (from the "other" factor matrices)
  - Khatri-Rao product (KRP) of d factor matrices
  - $N = O(n^d)$, i.e. very large
  - Structure of this matrix is key!
- Matrix $\mathbf{X}^{\mathsf{T}}$ (from the data tensor)
  - Transpose of mode-unfolding of tensor
  - Sparse if tensor is sparse, including all-zero rows
  - Cannot afford any transform that destroys sparsity
- Matrix $\mathbf{A}^{\mathsf{T}}$ (transpose of factor matrix)

CP-ARLS Leverage Applies Sketching to CP Least Squares Problem

How can we utilize sketching in the inner least squares problem of CP-ALS?



$$\min_{\mathbf{A}} \|\mathbf{S}\mathbf{Z}\mathbf{A}^\top - \mathbf{S}\mathbf{X}^\top\|^2$$



Corresponds to sampling fibers from the tensor

Battaglino, Balladrd, and Kolda (2017). "A Practical Randomized CP Tensor Decomposition"

Ingredient #1: Leverage Scores Key to Limiting Samples (But too Expensive to Compute)

Leverage score: $\mathbf{Z} \in \mathbb{R}^{N \times r}$

"Economy" SVD:

$$\mathbf{Z} = \boxed{\mathbf{U}}\mathbf{\Sigma}\mathbf{V}^T$$

Spans column space of $\mathbf{Z}$

Leverage score of row i:

$$\ell_i(\mathbf{Z}) = \|\mathbf{U}(i,:)\|_2^2 \in [0,1]$$

Leverage scores indicate which rows most influence solution but cost $O(Nr^2)$ to compute

Rough Intuition:

$$\ell_1(\mathbf{Z}) = 1$$

$$\sum \ell_i(\mathbf{Z}) = r$$

How "important" a row is to column space.

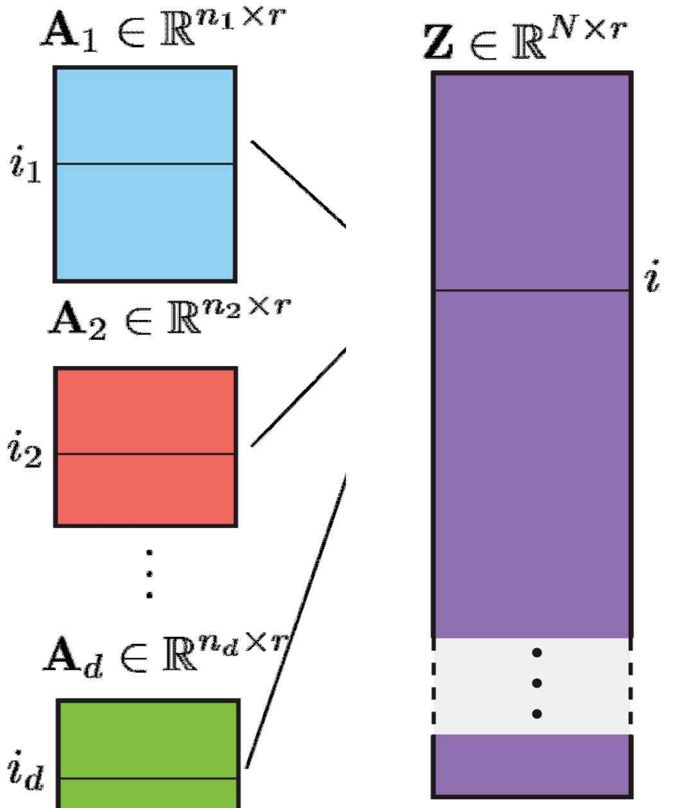
$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1r} \\ 0 & z_{22} & \cdots & z_{2r} \\ 0 & z_{32} & \cdots & z_{3r} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & z_{N2} & \cdots & z_{Nr} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \vdots \\ \nu_N \end{bmatrix}$$

Can use approximate leverage score, pay penalty in required sample number

Survey: D. P. Woodruff, ***Sketching as a Tool for Numerical Linear Algebra***, 2014

Ingredient #2: Exploit KRP Structure to Bound Leverage Scores

$$\mathbf{Z} = \mathbf{A}_d \odot \cdots \odot \mathbf{A}_1$$



$$N = \prod_{k=1}^{d} n_k$$

Upper Bound on Leverage Score

Lemma (Cheng et al., NIPS 2016; Battaglino et al., SIMAX 2018):

$$\ell_i(\mathbf{Z}) \leq \prod_{k=1}^{d} \ell_{i_k}(\mathbf{A}_k)$$

Too expensive to calculate $O(Nr^2)$

Cheap to calculate individual leverage scores $O(r^2 \sum_k n_k)$

Set probability of sampling row i to:

$$p_i \equiv \frac{1}{r^d} \prod_{k=1}^{d} \ell_{i_k}(\mathbf{A}_k)$$

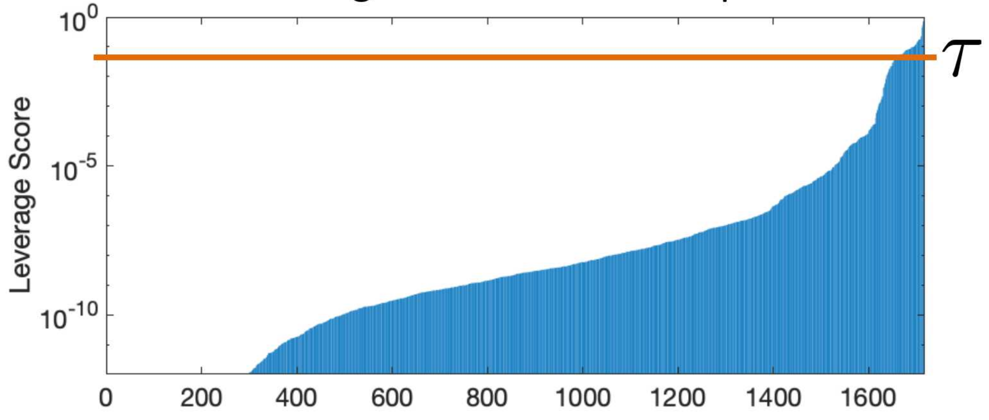
Sample row from each factor matrix to avoid forming all N possible combinations corresponding to rows of $\mathbf{Z}$!

1-1 Correspondence between *linear index* and *multi index:*

$$i \in [N] \Leftrightarrow (i_1, \ldots, i_d) \in [n_1] \otimes \cdots \otimes [n_d]$$

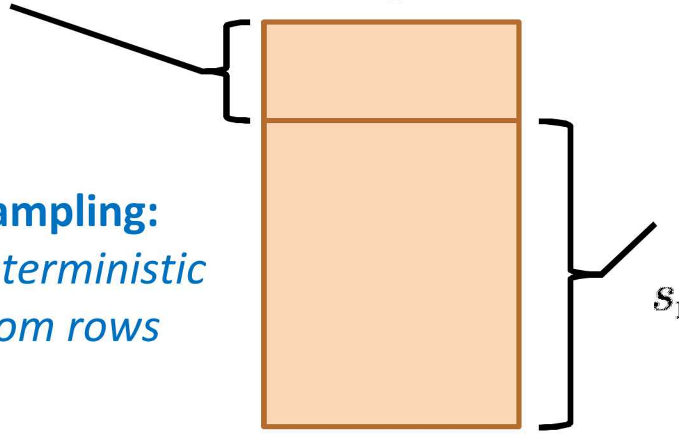
Ingredient #3: Deterministically Include All High-Probability Rows

Sorted Leverage Scores from Real Sparse Data



Deterministic Rows Above τ

$\mathbf{SZ} \in \mathbb{R}^{s \times r}$

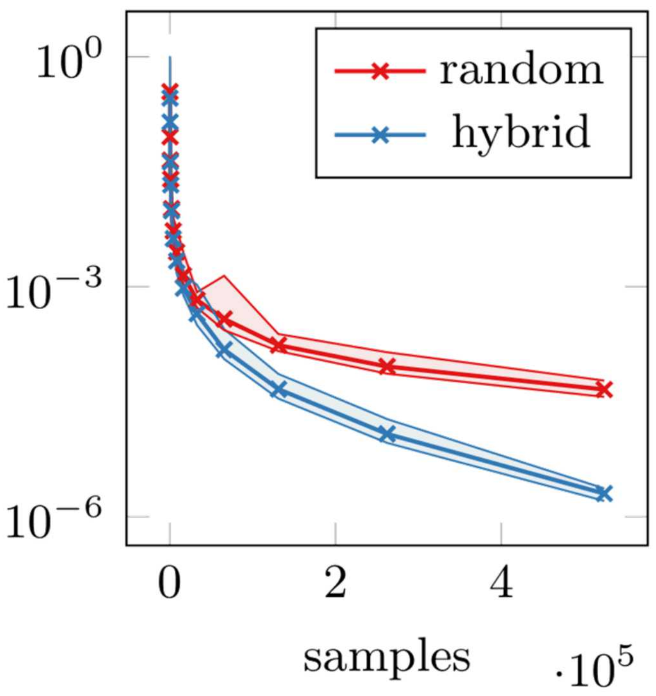


Random Rows

$s_{\text{rnd}} = s - s_{\text{det}}$

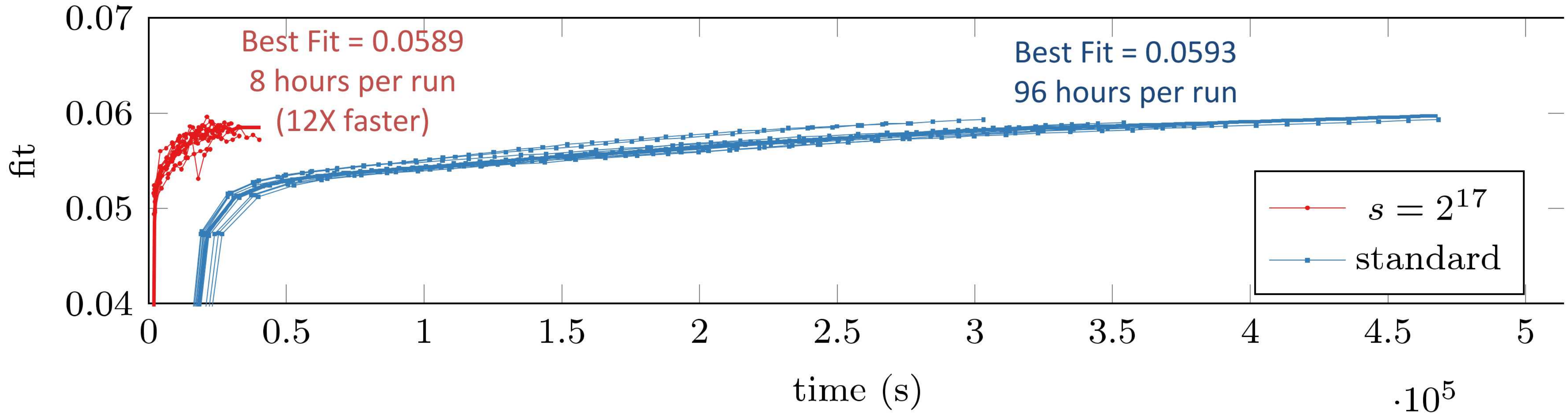
Hybrid Sampling:
Combine deterministic and random rows

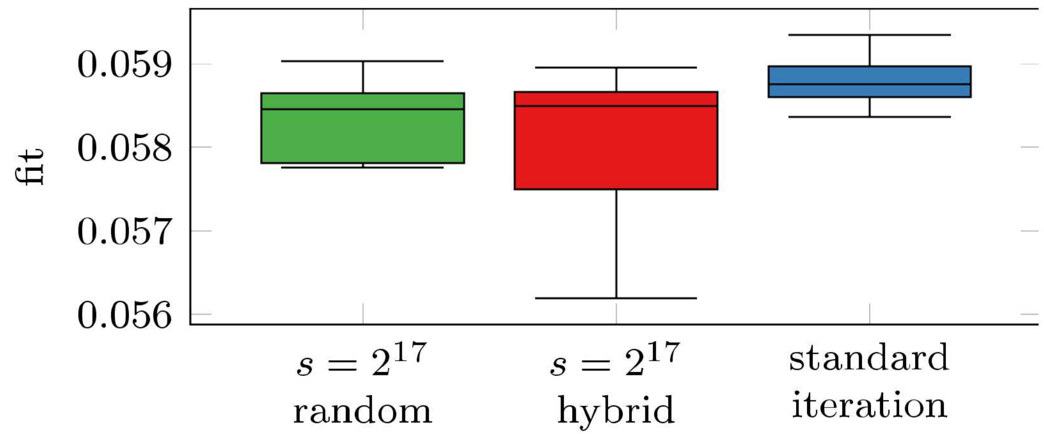
Difference to True Residual



Single Least Squares Problem with N = 46M rows, r = 10 columns, n = 183 right-hand sides

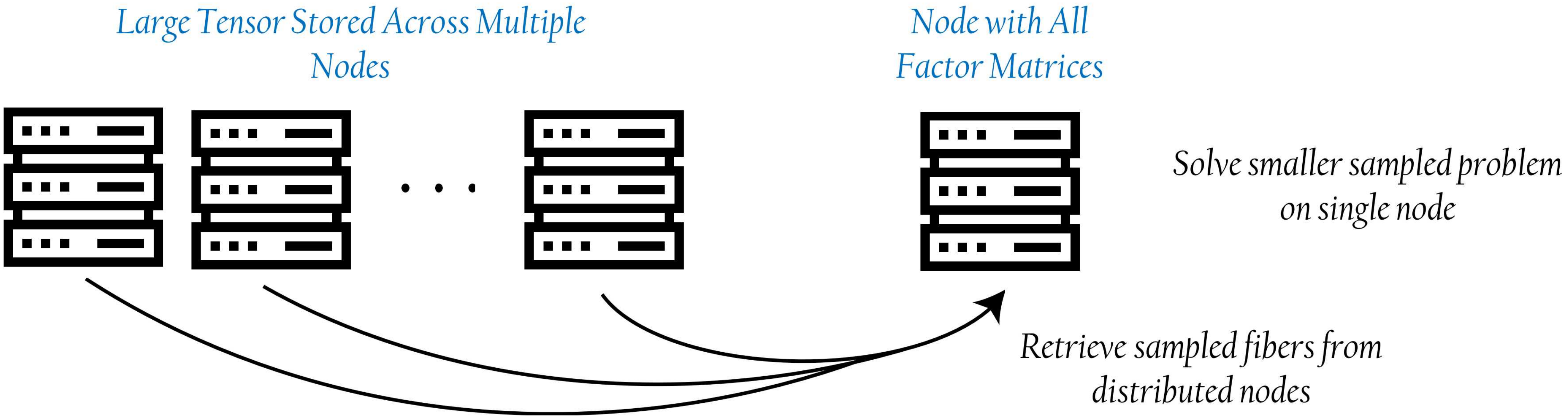
Numerical Results on Reddit Tensor





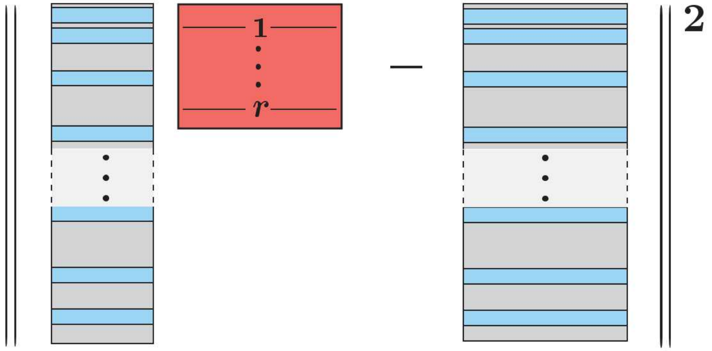
# Samples (S)	Speedup
2^{17} Random	16.08
2^{17} Hybrid	11.87
CP-ALS	1.00

Extensions to Distributed Data and Summary



<ins>**Summary**: Sketching for Sparse Tensor Problems</ins>

1. Sampling by leverage scores minimizes required samples
2. KRP structure enables efficient leverage score estimates
3. Hybrid sampling for concentrated leverage scores



Questions

B.W. Larsen and T.G. Kolda. *"Practical Leverage-Based Sampling for Low-Rank Tensor Decomposition"* on arXiv.